



LIBRARY
OF THE
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

LD28
.M414
no. 520-
71
c.2

MASS. INST. TECH.
MAY 10 1971
LIBRARIES

WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

Research Program on the Management of
Science and Technology
The Delphi Technique: Survey and Comment

520-71 Alan R. Fusfeld* March, 1971

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139



Research Program on the Management of
Science and Technology
The Delphi Technique: Survey and Comment

520-71 Alan R. Fusfeld* March, 1971

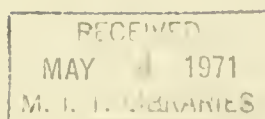
* Alan R. Fusfeld is a doctoral candidate at the Alfred P. Sloan
of Management, The Massachusetts Institute of Technology, Cambridge,
Massachusetts.

HD28

.M414

no. 520-71

C.2



THE DELPHI TECHNIQUE: SURVEY AND COMMENT

by Alan R. Fushfeld

Even though DELPHI has been used with some success, ..., it should not be interpreted as a device that produces 'truth about the future'. The Delphi method is designed to produce consensus judgments in 'inexact' fields; it would be a mistake to consider such judgments as complete or precise descriptions about the future.

Theodore J. Gordon and Robert Ament,
"Forecasts of Some Technological and
Scientific Developments and their
Societal Consequences", IFF R-6,
September, 1969

Introduction:

Delphi forecasting is a widely used technique for the systematic development of expert opinion consensus concerning the future (1). This article considers the implications for Delphi of recent analyses of forecast results and of the historical development of the Delphi methodology.

It is important to note at the outset that the theoretical utility of Delphi is that it pools individual opinions in order to develop forecasts which are more accurate than those of any average individual. This has been the major driving force behind its development.

The potential utility of Delphi, described above, was recognized by Norman Dalkey and Olaf Helmer in 1951-1952. They were working for the RAND Corporation and applied Delphi methodology to Air Force strategy problems in Project "DELPHI" (2). Their work, however, was not oriented towards the timing or nature of future events, as we speak of these characteristics today. Rather, they and other early developers were interested in

policy formation research, operations research, and political gaming (3). Some of these early uses were directed at: (1) estimating Soviet bombing schedules for U.S. locations, in the case of war on July 1, 1953; (2) predicting election outcomes; and (3) predicting U.S. policy towards Communist China. Attempts were also made to validate the methodology through "almanac" delphi forecasts, which are predictions made for questions with obscure but known answers.

This early work was extended in 1963-64, by Ted Gordon working in conjunction with Olaf Helmer, to include the long range forecasts, which are now familiar. Since then, Delphi studies have been conducted in such diverse fields as medicine, education, space hardware, oceanography, and transportation.* The sponsoring organizations have also been widely varied-- government agencies (U.S. and foreign), RAND, IFF, Goodyear Tire and Rubber Corp., TRW Corp., LTV Corp., McGraw-Hill Corp., et. al. (1, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19).

A brief consideration of the summary description of the development of Delphi, just presented, prompts several interesting questions. The two which are most important, because of their vital positions in the rationale of Delphi methodology are:

- (1) how valid are the "almanac" delphi forecasts?
- (2) how similar are long range delphi forecasts to the processes of estimation and short-term prediction, for which Delphi was first devised?

An understanding of these questions is essential to the efficient utilization of Delphi forecasts in corporate planning.

The remainder of this article is structured into the following five sections to help provide the understanding:

- (1) background and basic rationales for the development of Delphi forecasting;
- (2) extensions to Delphi methodology;
- (3) real-world uses of Delphi (includes cases, which illustrate uses of material presented in the preceding sections and provides a focus for questions concerning the behavior of Delphi);
- (4) recent studies and implications of forecast results; and
- (5) corporate use of Delphi (draws preceding material together to suggest how corporations might increase their benefits from Delphi, through an improved understanding of its nature).

*The Institute for the Future (IFF), Riverview Center, Middletown, Conn., perhaps, has the greatest amount of available information regarding current Delphi studies (over 2000). Olaf Helmer is president of IFF and Ted Gordon is a senior fellow.

Background and Rationales of Methodology:

The background of Delphi methodology extends into the philosophy of the solicitation of experts' opinions. Norman Dalky and Nicholas Rescher give a fine description of this and contend that there are certain fields of knowledge, which, by their nature, commonly require the opinions of an expert (20). These fields are termed "inexact," meaning that absolute answers are not known or possible, and include such areas as jurisprudence, political policy formation and corporate strategy.*

The opinions of experts, then, are justified as inputs to decision making, where judgment ^{is required} in an "inexact" area. Of course, expert judgment has been sought by decision makers for thousands of years, but, unfortunately for many of the experts, the judgments were not always correct.** Some striking examples of misjudgments by experts, in the "inexact" area of future technology, appear below:

(1) "... The bow is a simple weapon, firearms are very complicated things which get out of order in many ways... and tires out soldiers on the march. Whereas also a Bowman can let off six aimed shots a minute, a musketeer can discharge but one in two minutes." (Advice to the British Privy Council in 1591 by Colonel Sir John Smyth.) (21)

(2) "There is no plea which will justify the use of high-tension and alternating currents, either in a scientific or a commercial sense. They are employed solely to reduce investment in copper wire and real estate." (Edison, T.A., "The Dangers of Electric Lighting," North American Review, November, 1889) (21).

(3) "... We hope that Professor Langley will not put his substantial greatness as a scientist in further peril by continuing to waste his time, and the money involved, in further airship experiments. Life is short, and he is capable ^{of} services to humanity greater than can be expected to result from trying to fly...." (New York Times editorial, December 10, 1903... one week before the first flight of the Wright brothers) (21).

* The opposite of "inexact" fields would be "exact" or precise areas of knowledge. Examples of "exact" sciences are physics, mathematics, and chemistry. It should be clear that many areas, particularly the social sciences, combine "inexact" and "exact" sub-fields.

** The reader, who is not already familiar with the origins of the word "DELPHI," is reminded that it refers to the oracle of Apollo located at Delphi, a place in ancient Greece (22).

Given the fact that single experts may hold incorrect opinions, it is logical to seek increased accuracy through the opinions of a collection of experts. This assumes, though, that a representative statistic of the experts' opinions may be chosen, which will be more accurate than the opinion of the average expert, i.e. it will somehow be corrected for individual bias and misinformation.

One of the first formal attempts to systematically aggregate the opinions of experts was "Project DELPHI," begun at RAND Corporation in 1951-1952. The work was conducted by Norman Dalkey and Olaf Helmer, who hoped to develop a methodology for combining the opinions of experts on the subject of estimating Soviet offensive strategies (2).

Dalkey and Helmer built their study on earlier work in the area of public opinion research. * A 1948 study in this area, "The Prediction of Social and Technological Events," by Kaplan, Skogstad, and Girshick, is particularly interesting, since it: (a) illustrates the background in which "Project DELPHI" developed and (b) presents evidence that statistical aggregates of individuals' opinions, for near term events, are more accurate than predictors derived from group interaction or single individuals (3).

The Delphi technique was formed by adding the concepts of controlled information feedback and iterations of responses to the aggregation of individual opinion. An additional aspect of the technique eliminated some psychological effects of groups (e.g. fear of superiors, fear of changing opinion) by keeping the individuals anonymous to each other. Delphi methodology provided estimates for the RAND policy/strategy studies, that should have been more accurate than those of other types of groups or of individuals for two reasons: (1) the avoidance of psychological biases of committees, panels, etc.; and (2) the increase in accuracy from a pooling of expert judgment.

Following the development of the Delphi technique, attention was turned towards experiments designed to test its validity (13,14,15). These experiments applied the Delphi methodology to questions where the answers were known but obscure (e.g. How many tons of potatoes were harvested in Idaho in 1960?). The responses to such 'almanac' questions were shown to:

(1) converge during several iterations; and

(2) converge towards the correct answer for the majority of the questions.**

It is the "almanac" Delphi results which were and are used to verify the usefulness of Delphi.

* The major journal of this field is the Public Opinion Quarterly, established in January, 1937.

** The analysis is discussed in detail in a later section.

As indicated earlier, the Delphi methodology was not generally applied to questions concerning the future until 1963-64. Several things happened which influenced this. First was the release for publication of some of the "Project Delphi" work in 1962. (The subject matter of the work has required classification for security reasons.) In addition, there was a growing interest at this time in forecasting technological developments, i.e. an increased awareness of the impacts of technological change on long-range plans and an increased demand for quantification in all aspects of planning (e.g. PPBS). Finally, in 1963, Ted Gordon (Director of Space Station and Planetary Systems for Douglas Aircraft Company) met Olaf Helmer by chance, during a luncheon meeting at RAND (5). It was at this point that Gordon's interests in technology planning were linked to Helmer's interests in the utilization of expert opinion. A consequence of this was the application of the Delphi technique to future-oriented questions.

The Helmer-Gordon work was the first use of Delphi for long-range (greater than one year) forecasting (17). Their study ("Report on a Long -Range Forecasting Study") concentrated on six areas of interest: (1) scientific breakthroughs; (2) automation; (3) space; (4) population; (5) war prevention; and (6) weapon systems. These were different in two respects from the topics considered in "Project Delphi." First, Helmer-Gordon topics required an estimation of all of the many interacting dynamics of actors influencing technological advance, whereas "Project Delphi" topics required an estimation of strategies based on role playing, with technology and other variables fixed or in a static state. Also, the relative time horizons of the topics were different, i.e. a year or less for "Project Delphi" and unlimited time for the long-range study. (These differences are important for the comparison and evaluation of future oriented studies, "almanac" studies, and "Project Delphi" studies, discussed in detail in a later section.) It is clear , though, that regardless of differences between the two applications of Delphi methodology, the Helmer-Gordon work marked the beginning of the use of the Delphi technique for technological forecasting.

Delphi methodology, then, became a technique for forecasting technology, through the linking of quantitative planning needs with the procedures developed by "Project Delphi." The theory and rationales also come from this background (5, 6, 10, 2, 20, 3, 23, 26, 17). They may be summarized by

the following structure, consisting of a major premise, several minor premises, and a group of postulates:

Major Premise:

Future events occur due to the complex interactions of many forces.

Minor Premises:

1. Many elements involved in the projection of future developments are not easily quantifiable.
2. In non-exact disciplines, expert opinion "modeling" must, of necessity, substitute for the exact laws of causality found in the physical sciences.
3. The prediction of outcomes or occurrences of future events, with any degree of probability, is an "inexact" science, just as jurisprudence or organizational strategy is.
4. Individual experts suffer biases and are incorrect nearly as often as they are correct in predicting future events.

Postulates:

1. It is justifiable to solicit and bring to bear expert knowledge and judgment concerning future developments.
2. Groups of experts have aggregate-relevant information that is equal to or greater than any member of the group.
3. Groups of experts have aggregate misinformation that is equal to or less than any member.
4. Groups of experts have at least as many informal models (systematic processes for analyzing information), by which an answer is estimated, as any member.
5. Groups which meet face-to-face, in order to develop a consensus, suffer bias from psychological pressures (e.g., follow the leader syndrome or reluctance to change previously-stated opinion).
6. Groups whose members are anonymous to each other, develop a consensus which corrects for an individual bias and removes the possibility of a psychologically-developed bias.

7. The median response of a group is at least as good as one half of the respondents (by definition).

Together, the postulates imply that the application of the typical Delphi technique, consisting of

- a) anonymity of experts - to eliminate bias,
- b) controlled feedback - to develop consensus, and
- c) an estimator of group opinion, will provide a more accurate response to a question than that of the average expert.

A description of the typical Delphi process appears below and is accompanied in Figure 1 by several examples of Delphi forecasts. It should be apparent that the actual process fits in well with the preceding discussion of the underlying premises.

A typical [forecast] is initiated by a questionnaire which requests estimates of a set of numerical quantities, e. g. dates at which technological possibilities will be realized, or probabilities of realization by given dates, levels of performance, and the like. The results of the first round will be summarized, e. g. as the median and inter-quartile range of the responses, and fed back with a request to revise the first estimates where appropriate. On succeeding rounds, those individuals whose answers deviate markedly from the median (e. g., outside the inter-quartile range) are requested to justify their estimates. These justifications are summarized, fed back, and counter-agreements elicited. The counter agreements are in turn fed back and additional reappraisals collected.¹⁰⁾

[Preliminary questionnaires are often used to select and develop questions for which estimates will be collected.]

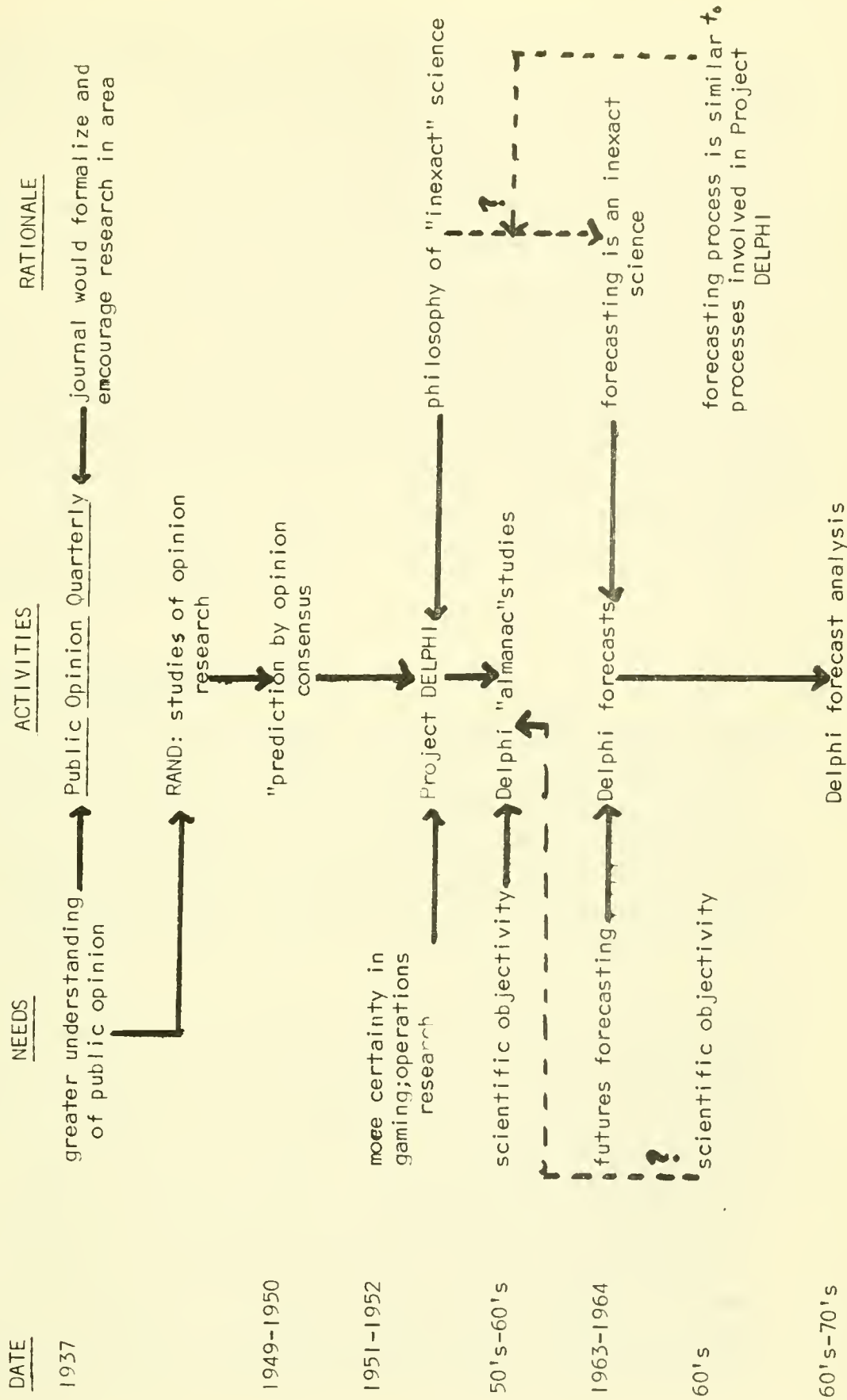
In summary, the development of the Delphi technique to 1964, as a new tool for forecasting, followed a logical sequence of steps. This is illustrated in Figure 2 which also serves to highlight two aspects of the development which should be of particular interest to present and

Figure 1:

EXAMPLES OF DELPHI FORECASTS

Description of Event	Date By Which			Year Forecast Made	Reference
	25	50	75 %		
	of Respondents Think Event Will Occur				
There will be a single national building code.	1975	1977	1980	1968	18
Polymers will be created by molecular tailoring, with service temper- ature ranges in excess of 1000 F.	1971	1976	2000	1968	18
SST aircraft will be in regular service over land areas.	1980	1982	1985	1968	8
Hydrocarbon/air fuel cell will be commercially available.	1974	1983	1990	1967	27
A source of transplant organs for humans will be developed through selective breeding of animals that are tissue compatible.	1990	2015	2015	1969	11

DEVELOPMENT OF DELPHI FOR TECHNOLOGICAL FORECASTING



potential users of Delphi (see dashed lines in Figure 2). The first is the tacit assumption that forecasting is similar to the processes of estimation involved in Project DELPHI and the judgmental processes that are justified by the philosophy of the inexact sciences. This is not intuitively obvious and may result in overjustifying the use of Delphi for future technological developments, which, in turn, may create a sense of overconfidence in the forecast results. The other aspect of importance is the establishment of scientific objectivity for Delphi forecasts through "almanac" studies. Given that almanac studies do, in fact, justify the use of Delphi for almanac studies it is again not intuitively obvious that they are adequate tests for forecast studies. (A discussion of the implications of almanac studies' results is in a later section.) It is suggested that users of Delphi forecasts be aware of these tacit assumptions and their implications and should, perhaps, exercise caution in placing too much confidence in the results.

2. Extensions to Delphi Methodology

After 1964, several additions were made to the Delphi technique. None of them altered its basic design, but rather they were aids to increase its utility or efficiency. These extensions consist of (1) cross-impact analysis; (2) SEER (System for Event Evaluation and Review); and (3) integration of trend extrapolation with Delphi^(3, 4, 6, 19). Each is discussed briefly below and, of course, is based on the underlying presumption that Delphi is an acceptable, if not perfectly accurate, forecasting technique.

Cross impact was developed to solve the problem of inter-relationships between forecasted events (28, 29). In other words, forecasts which consist of collections of potential future occurrences, likely dates of their occurrence and their probability might contain mutually reinforcing or mutually exclusive events. An example of a reinforcing inter-relationship would be the discovery of the means for the magnetic containment of magneto-hydrodynamic plasmas positively affecting the development of commercial thermo-nuclear power. A negative impact is illustrated by the effect of continued inexpensive sources of coal on the development of economically viable nuclear power. Basically, the

technique consists of establishing a matrix of the cross-effects of the various items in the forecast, such that a consistent set of estimates can eventually be developed.

The evaluation of cross-effects is the key variable in cross-impact analysis. Usually cross-effects are determined by consensus and may be in the form of a revised conditional probability estimate for an event or in the form of a general type of influence that the occurrence of one event has upon the probability of occurrence of another.* Consistency of forecast results is achieved through a series of iterations which revise the evaluations of the cross-impact matrix. When simplifying assumptions are made, this can be done by computer. Eventually, a complete matrix is developed which indicates modes of linkage between events, strengths of relationships, and predecessor-successor orientations.

Thus, cross-impact analysis is an important technique for the user of Delphi, since it eliminates contradiction and derives an internally consistent forecast of revised probabilities of events, at least some of which are interrelated.

Another addition to Delphi is SEER -- System for Event Evaluation and Review(4). This technique helps to make Delphi forecasting a more efficient process by developing initial lists of events through interviews, thereby avoiding a flow of many forecasts which start with a blank page and by clearly indicating a set of rules, which reduces the amount of time required by the Delphi "panelist". The time needed by the forecast process can be quite considerable, particularly when hundreds of events are involved. SEER limits the time investment by constraining the forecast to two rounds and requesting that panelists answer only those questions within their areas of expertise. SEER also addresses the question of "event desirability" from a user's point of view and "event feasibility" from a producer's point of view. Finally, utilization of this technique produces a Delphi forecast in which interrelationships, goals, supporting events, and alternative paths are clearly specified.**

* Examples of general types of influence would be enhancing (enabling or provoking) and inhibiting (denigrating or antagonistic) [28].

** The interested reader is advised to see: "SEER: A Delphic Approach Applied to Information Processing", by G. B. Bernstein and Marvin Cetron, in Technological Forecasting, VI, N. 1, June 1969.

-10-

The third extension to the Delphi technique is an integration of cross-impact analysis, SEER, and trend extrapolation with the Delphi process (13, 16). The new factor in this approach is the application of Delphi to the determination of future trends. Panels are given the historic (time series) data and asked questions designed to elicit their opinions concerning the future environment and direction of trends. The design and results of this process are similar to SEER and cross-impact matrices. It is a significant addition to Delphi since it forces the experts to relate their intuitive models of the future to actual, as opposed to perceived, historical trends.

3. Case Applications

At this point, it is useful to consider brief discussions of how Delphi forecasts have been made and utilized by four different organizations. They indicate the range of complexity that has been attempted and include Ling-Temco-Vaught Corporation, TRW, Institute for the Future, and Goodyear Tire and Rubber Corporation.

LTV initiated technological forecasting (basically in terms of Delphi) in 1968 (8). Their forecasts consisted of approximately one hundred events and were incorporated into corporate plans through inclusion in their R&D planning cycles. Panelists with non-technical backgrounds were used to broaden the base of intuitive models and thus increase the awareness of the Delphi group to non-technical aspects of the questions, e. g. economic, marketing, or environmental factors. In addition, the feedback for iterations included the percent of "never" responses in order to preclude the possibility of trends toward conformity which were observed in certain LTV tests without the percent of "never's" feedback. Finally, it should be noted that the resulting forecasts were made with estimates of probability and indications of which LTV divisions would be primarily effected by each event.

Complex but straight Delphi forecasts are also the trademark of TRW's forecasting efforts which began in 1966-67 (6, 7). Their studies have covered diverse fields -- space, military, transportation, housing, communication/education, materials, power, and biology/oceanography and contained hundreds of events. TRW integrated their forecast information through the use of future "mapping" in order to

enhance the utility of such diverse fields of information for corporate planning. A succinct description TRW's mapping is given below by Donald L. Pyke, Manager, Technical Liaison and Forecasting, TRW, Inc.:

...[Mapping is] a concept for a comprehensive system which utilizes available forecasting techniques to generate a map of the technological future which will provide for the planner or assessor a tool similar to that available to the travelers. As envisioned, this map would display technological alternatives and their environmental consequences in such a way as to enable the coupling of near term technical activities to long range forecasts... For maps of the future, one axis of the coordinate system is time and the other is the category under investigation. An hierarchical system is envisioned in which the summary level of the environment subdivided into the eight aspects identified... [as] technological, commercial, political, social, ethical, cultural, personal and ecological... (7)

It is clear that TRW has designed a sophisticated vehicle for its use of Delphi forecasts, that, perhaps in simpler form, should interest many others.

An organization that contrasts with LTV and TRW is the Institute for the Future, which is concerned primarily with contract forecasting studies. In one of their many Delphi studies, they combined traditional event forecasting with Delphi forecasts of trends, panels for role playing, and cross impact analysis (13,14,15). This is probably the most sophisticated use of Delphi forecasting. It is particularly interesting to note the use of panels for role playing. These were employed to simulate the attitudes of various sections of society at various times in the future, given alternative sets of 'facts'. Panels, such as these, might also be very useful for corporate planners.

In order that the reader not think that all Delphi studies must be as complex and sophisticated as those discussed above, the following application by Goodyear is presented (9,30). They went through three simple steps. First, a group of experts from the principal organizations which influence the tire industry were requested by mail "...to list the most important developments they anticipate between now and the year 2000" (30). * Next, a relatively few event questions of particular importance were selected. Several questionnaire iterations of the revised list

*The organizations included tire companies, fiber companies, chemical companies, carbon block companies, industry consultants, and technical trade journals.

concluded the study. It is important to note, that this study, which is less complicated than the others mentioned earlier, was judged as very useful to Goodyear's research plans.

The reader is probably aware that the forecasts just discussed are based upon an acceptance of the accuracy of Delphi. However, this is a tacit assumption that has not been clearly explored, either, by the rationale of Delphi methodology or by empirical evidence because of the long range nature of the forecasts. Thus, until confirming evidence is available, the planner who intends to use the Delphi process as a method for aggregating intuitive judgements concerning the future, must, himself, decide two things: (a) how accurate does he believe Delphi forecasting is? and (b) to what extent can Delphi forecasting be relied upon--for what length of time? If this is not done, even relatively simple techniques may lead him, unaware, into future difficulty.

4. Behavior and Analysis of Delphi

In the last several years, a number of studies have been conducted on Delphi 'almanac' data and Delphi forecast data (24, 25, 31, 32, 33, 34, 35). An analysis of these studies provides additional information for understanding the behavior of the Delphi process. A summary of this analysis and the implications for the use of Delphi are presented below in the following framework:

1. behavior of Delphi responses for 'almanac' questions;
2. behavior of Delphi responses for future questions;
3. summary of analysis of results of the 1964 Delphi forecast; and
4. other comments.

The reader will recall that Delphi 'almanac' studies consist of responses to questions with known, but obscure answers. Some actual sample questions are: "--how many million board feet of lumber were produced in the United States in 1962? --what was the average price received by the United States farmer for a bushel of apples in 1940? --what is the specified operational gross weight, in pounds, of the Gemini capsule (exclusive of occupants)?" (23).

Such questions have been given to RAND research staff, graduate business students, and members of industrial corporations. The results have been quite similar and yield a number of interesting conclusions.

With regard to the accuracy of responses, a primary item of interest is that average group error decreases with (a) increases in group size, (b) increases in self-rating, and (c) decreases in the amount of dispersions among the respondents. This can be seen in Figures 3 and 4 quite clearly. A note should be made here that while dispersions and self-rating each correlate well with error when the other variable and group size (Figure 3) are held constant, there is no data to substantiate the independence of group size (Figure 4). In other words, with self-rating held constant, it is possible that the size of the group may decrease the group error by decreasing the dispersion (dispersion is dependent on group size) or the error may decrease with increased size for any given dispersion (dispersion independent of group size). Regardless, it is still of interest to note the effects of dispersion, self-rating, and, possibly, group size on the accuracy of response. These effects would appear to be useful in designing and evaluating Delphi studies.

Another area of importance in relation to the design of a Delphi is reproducibility of results. The 'almanac' studies indicate that results (medians) from different groups of the same size became more nearly alike as the number in the groups increases. The correlations between pairs of groups of various sizes are exhibited in Figure 5. This lends additional weight to the significance of group size, mentioned before with reference to accuracy, as a Delphi design variable.

The characteristics of 'almanac' studies discussed above are in reference ^{to} relative performance (accuracy, reproducibility) of different types of groups (differences in size, self-rating, and dispersion). These are important for the design of a Delphi study, but in order to maximize the utility of the Delphi it is necessary to go beyond relative performance to the performance behavior of a single group. There are three aspects to this:

1. the lognormal distribution of individual responses within a group;
2. the lognormal distribution of the median responses of the group about the true answers for a series of questions, and
3. the normal distribution of the relative accuracy of median responses to those of an individual in the group.

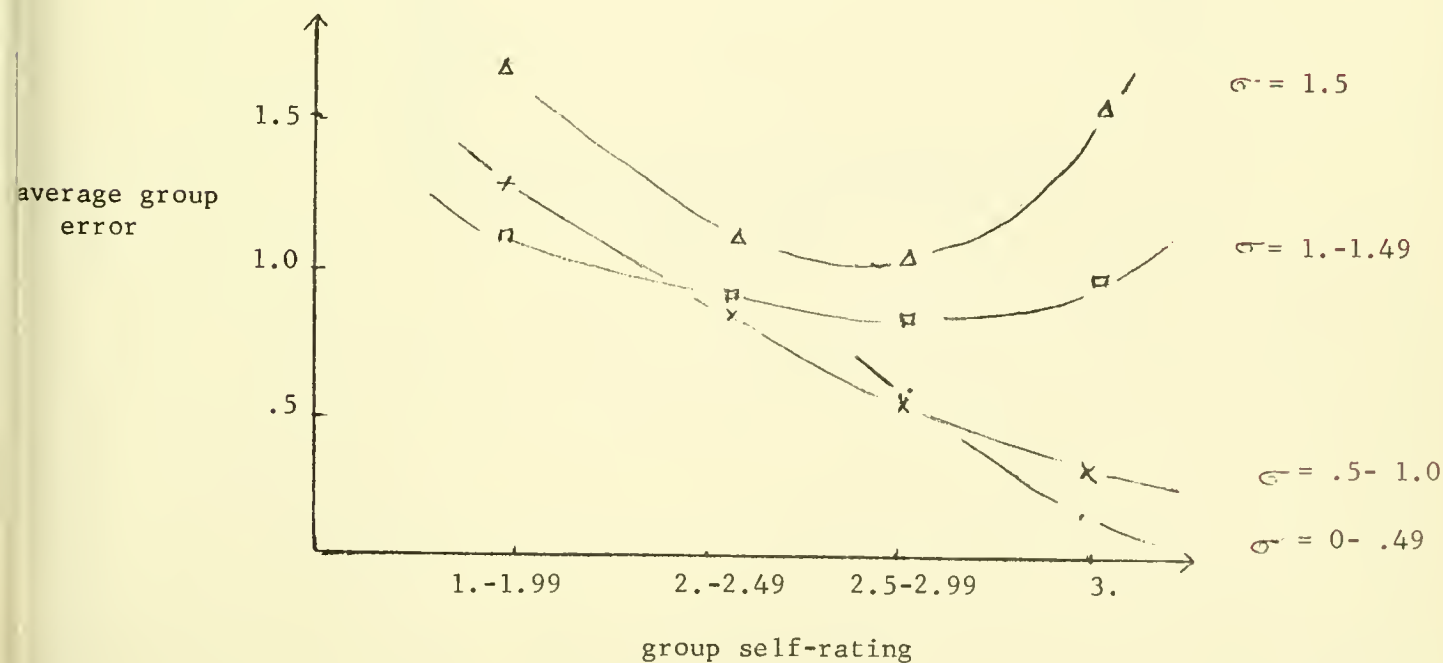
* Self-ranking refers to the individual respondent's indication of the degree of his familiarity with the subject material of a given question.

GROUP ERROR AS A FUNCTION OF
DISPERSION OF RESPONSES (σ)
AND GROUP SELF-RATING

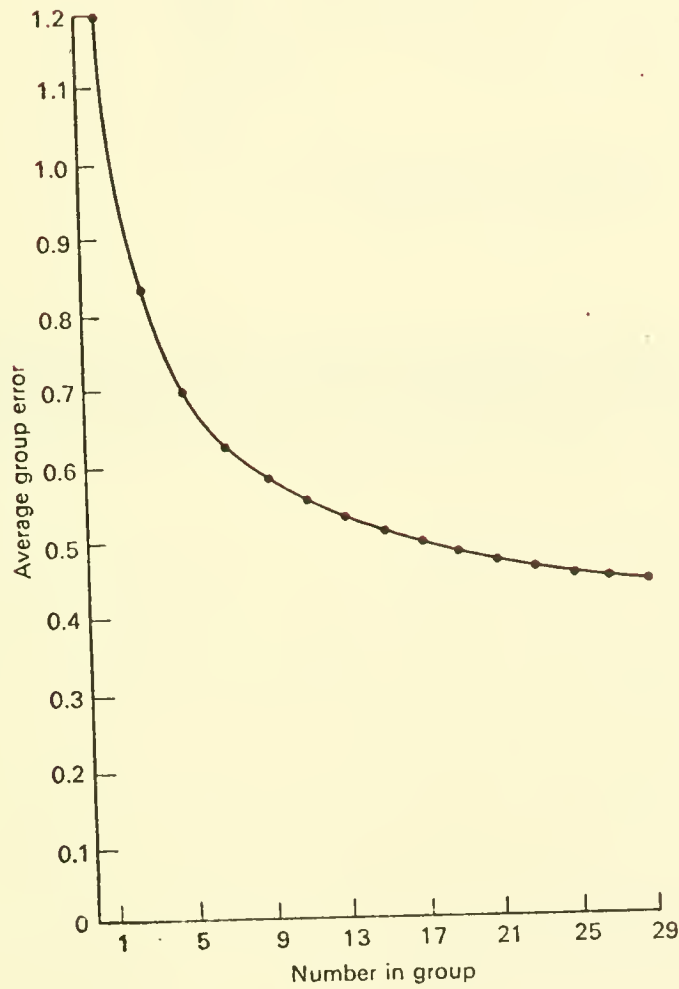
Round 1 Group self-rating \ σ	0-0.49	0.50-0.99	1.00-1.49	1.5 up
1-1.99		1.386 1	1.114 11	1.706 26
2-2.49		0.787 4	0.843 14	1.106 36
2.5-2.99	0.655 1	0.651 9	0.767 9	1.083 22
3 up	0.139 15	0.339 14	0.966 10	1.578 8

Note: all groups
have approximately
seven members

"For a fixed standard deviation, accuracy increases with group rating, and for a fixed group rating, accuracy decreases with increasing standard deviation. The anomalies in the lower right hand boxes may be accounted for by thin statistics; the number of cases is indicated in the small interior boxes." (25)

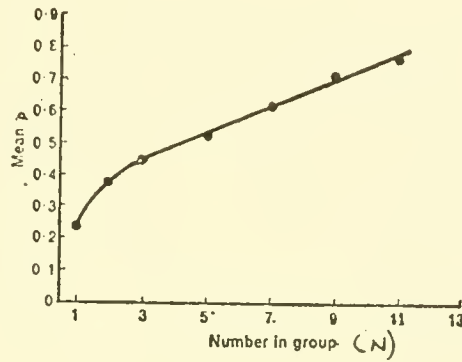


GROUP ERROR AS A FUNCTION OF
NUMBER IN GROUP



REPRODUCIBILITY OF RESULTS

$\bar{p}$ = mean coefficient of correlation between pairs of results of groups of size N



" It is clear that there is a definite and monotonic increase in the reliability of the group responses with increasing group size. It is not clear why the relationship would be approximately linear between $n = 3$ and $n = 11$. "

The lognormal distribution of the individual responses is illustrated in Figure 6. It is interesting because it indicates that the subjects in 'almanac' experiments develop their responses as ratios or relative magnitudes. This means, for example, that individuals select their answers from numbers such as 10, 100, and 1000, as opposed to 50, 100, and 150. It also means that if the correct answer is 100 and the responses are distributed about it in a random fashion, there will be an equal number of responses between 10 and 100 as between 100 and 1000. (There will not be an equal number between 50 and 100 and 100 and 150.) This implies that if individual answers fall equally to one side or another of the correct answer, the median will be a better estimation than the mean.

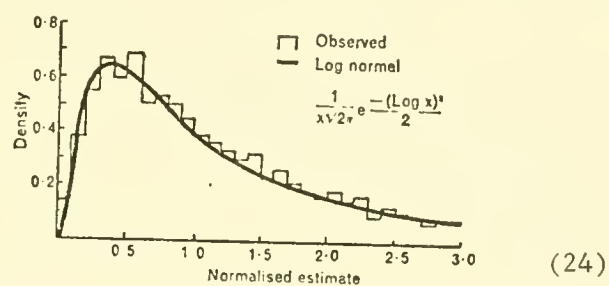
An analysis of median responses to a series of questions with respect to the correct answers also indicates a lognormal distribution (see Figure 7). This means that the distributions of answers of the entire group are randomly distributed about the correct answers. Further, they are distributed as ratios of the correct answers. The implication of this is that the panel, as a whole, may be off the right answer by a relative magnitude.

In addition to this, Figure 7 indicates (1) that most medians (i. e., most aggregate estimates by the group) tend towards the correct answer, (2) some medians may be magnitudes away from the correct answer, and (3) the dispersion of the medians tends to get smaller with more iterations. This is significant, since it should allow the development of confidence limits for the Delphi technique, as applied to 'almanac' question.

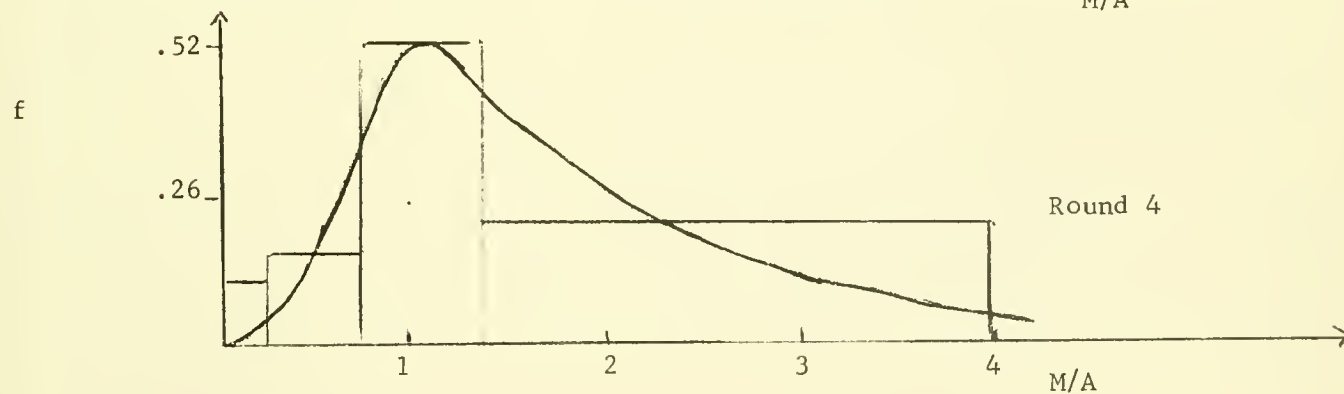
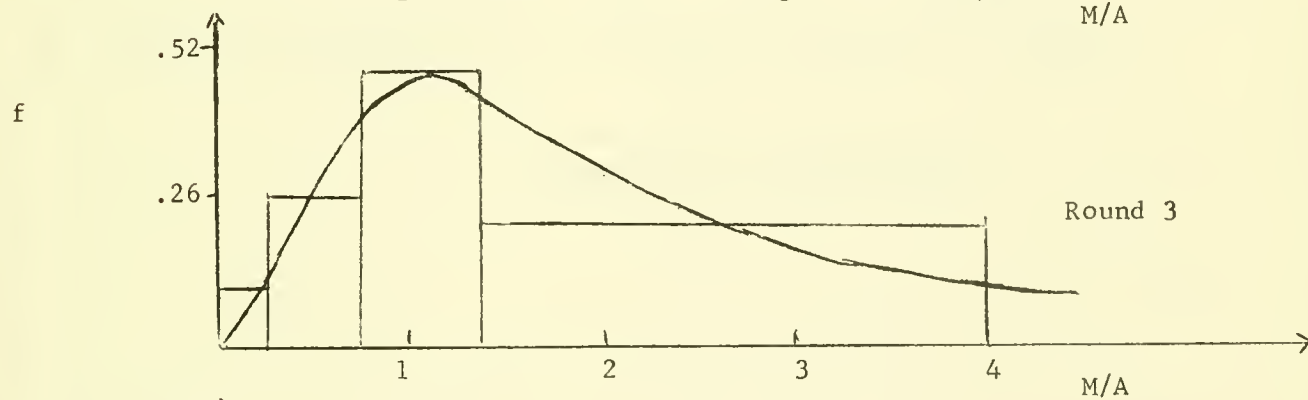
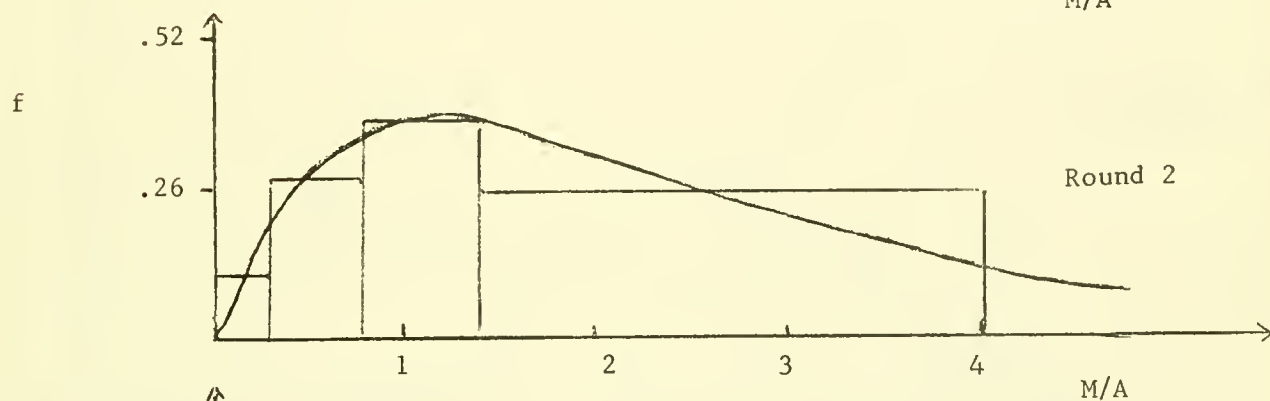
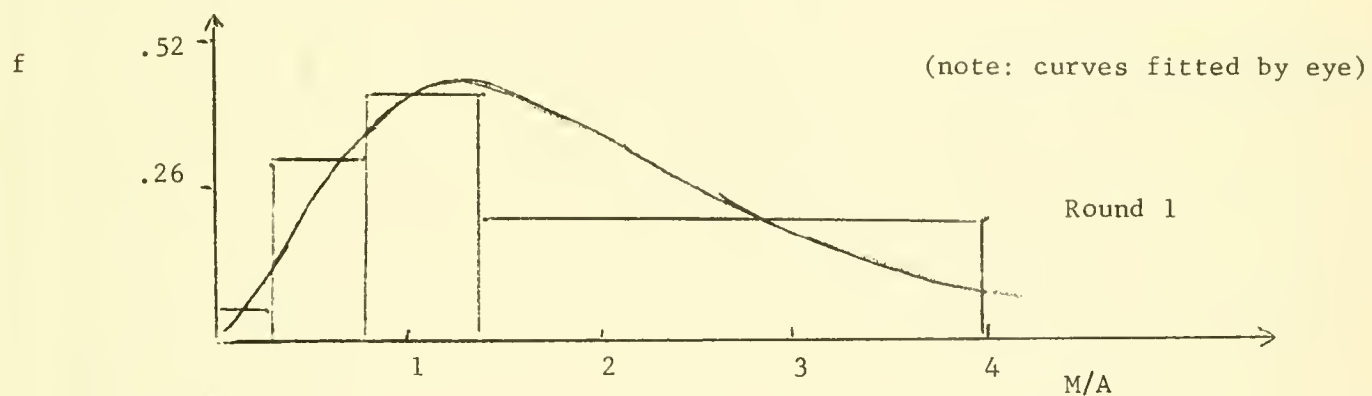
The 'almanac' Delphi also permits an evaluation of accuracy in terms of the median versus the individual respondent. For each individual, the number of questions for which the median was more accurate can be established. Of course, this number can vary from 0% to 100% of the question. The distribution of individuals over the various possible 'median-accuracy' categories is shown in Figure 8.

It would seem that some individuals always tend to do worse than the median, some always better, but most are more accurate for about half the questions. Put another way, the average individual's responses,

DISTRIBUTION OF INDIVIDUAL RESPONSES

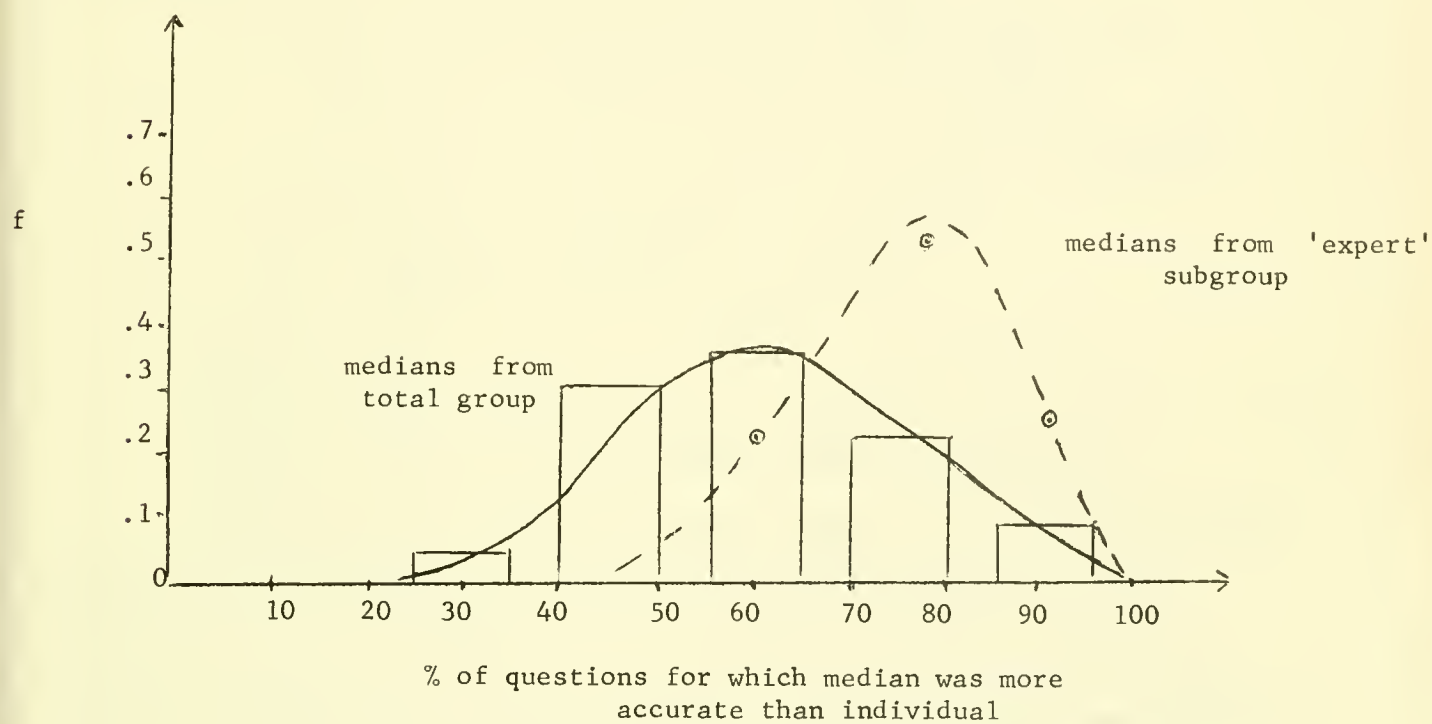


DISTRIBUTION OF MEDIANS (M) FROM A SERIES OF
QUESTIONS WITH RESPECT TO THE TRUE
ANSWERS (A)



DISTRIBUTION OF INDIVIDUALS WITH RESPECT TO
CATEGORIES OF RELATIVE MEDIAN IMPROVEMENT

(note: curve fitted by eye)



Data from ref. 23

if such a person exists, is as accurate or, perhaps, a shade less accurate than the medians.

This would seem to throw doubt on the usefulness of Delphi. However, when medians for each 'almanac' question are recalculated from a self-selected subgroup of individuals, they show an average improvement over the individuals' answers for 75% of the questions, as opposed to the previous 50% (see dashed line in Figure 9). The improvement demonstrates the usefulness of self-ratings for 'almanac' questions.

In summary, the results of 'almanac' Delphi studies show:

1. a dependence of relative accuracy on self-ratings;
2. a dependence of relative accuracy, on dispersion of replies;
3. a probable dependence of relative accuracy on group size;
4. a dependence of reproducibility on group size;
5. the lognormality of individual responses;
6. the lognormality of median responses about the true answers; and
7. the relative improvement over individual responses of medians taken from a self-selected subgroup of 'experts'.

In addition, the studies have shown Delphi is an improvement over face-to-face discussions, that consensus does develop with feedback, and that there is an optimum time (30 seconds) for reply to 'almanac' questions (24, 25, 32). Together, these results indicate important parameters for the design of a Delphi 'almanac' study and suggest that statistical analysis may be used to characterize its regular distribution behavior.

Recently, studies were completed indicating that results of Delphi forecasts exhibited similar behavior to that of the 'almanac' studies (32, 34). The two significant aspects of this work were the demonstration of the lognormality of Delphi forecasts and the growth dispersion of the respondents as the forecasts moved farther into the future.

Lognormal distributions of Delphi forecast respondents (see Figure 9) indicate that the panelists perceive the future in terms of relative magnitudes, just as the 'almanac' panelists perceived their answers. At the very best, this suggests that there is some underlying behavioral relationship that is common to processes of estimation. This is supported

LOGNORMALITY OF DELPHI FORECASTS

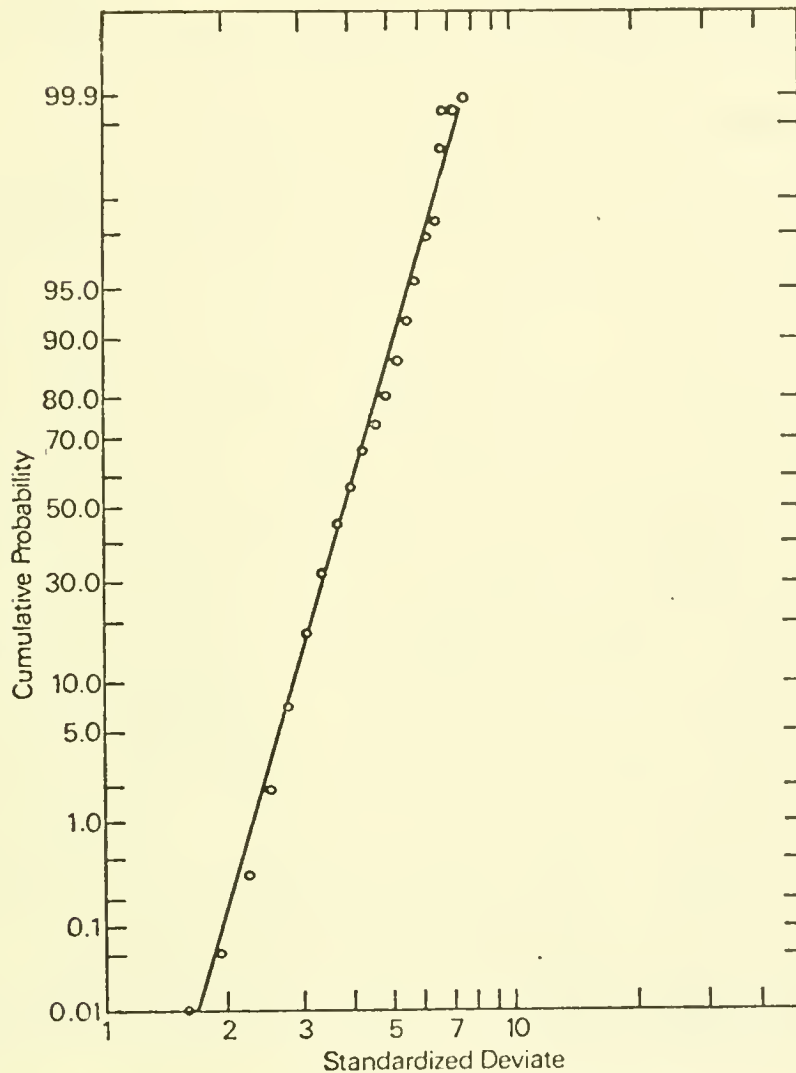


Figure 3. Cumulative probability distribution of standardized deviates for 90% likelihood estimates. A value of 4.0 was added to each deviate to make all results positive.

The data are plotted on lognormal coordinates. The good fit to a straight line indicates that the estimates are in fact lognormally distributed. It should be noted that to avoid the conceptual difficulties associated with plotting a "zero" on a logarithmic scale, an arbitrary positive constant was added to each of the deviates. This in no way affects the lognormality of the plot, and the original data can be recovered by subtracting the constant. The constant used depended on the range of values covered by the standardized deviates and is indicated with each figure.

This analysis of estimates obtained during an attempt to prepare a forecast corroborates earlier findings made through analysis of responses to almanac-type questions.

by the findings of several behavioral studies.* A recognition of the lognormality of the Delphi responses with respect to time is essential to estimating the amount of confidence that should be placed on a forecast as the time horizon increases and may help to explain the occurrence of median forecast dates that skip logarithmically into the future (i. e., forecasts regularly give median dates of individual years nearly to the year 2000 and then begin to skip 2005 - 2010 - 2025 - 2050 - later - never).**

The other important finding is the relationship between dispersion of respondents and median data of the forecast (33). It appears that the dispersion grows linearly with the time horizon at a rate that depends on the particular panel. This suggests that further attention should be drawn to factors which influence this rate. In addition, if one accepts the argument that Delphi forecast studies are at all like Delphi 'almanac' studies, then the increased dispersion should correlate with increased error. Hence, caution should be taken when using Delphi forecasts where the time horizon increases the dispersion beyond an acceptable level (judged, for the moment, individually).

At this point, it is appropriate to briefly mention the results of an analysis of Gordon's 1964 Delphi forecast (31). This study found that approximately 50% of the events were either accurately forecast or were forecast for the data by a 1969 forecast. It also showed the influence of environmental changes, since the 1964 space events, forecasted by technical agents, were made under the tacit assumption that funding would continue at 1964's level and/or growth rate. From the comfortable vantage point of hindsight we know that the funding environment has changed with obvious effects on the forecasts. It is clear from the preceding paragraphs that some forecasts are bound to be in 'left field' the planner should be aware

*The interested reader is advised to examine:

Stevens, S.S., On the psychophysical law, Psychol. Review, 64, 153-181, 1957.

Stevens, S.S., The psychophysics of sensory function, American Scientist, 48, 226-252, 1960.

Stevens, S.S. and E. H. Galanter, Ratio scales and category scales for a dozen perceptual continua, J. of Experimental Psychology, 54, 337-411, 1957.

Teichner, W.H., Psychophysical concepts of probability, Psychol. Reports, 10, 3-9, 1962.

**Note should be made here of the uncertainty estimation models of A. Wade Blackman (35). Essentially, he assumes that individual Delphi respondents have an underlying lognormal probability distribution relating to the event in question. He then applies Bayesian statistical techniques to revise forecast confidence limits.

of this. The 1964 study revisited shows some of the possible reasons for the variance in this case and, also, suggests some improvements, such as covering all of the 'bases', e.g., funding. Perhaps a 1964 Delphi should have been conducted with Congress. There is no guarantee that this would have helped, but it would have provided a more complete perceptual picture.

Other aspects of Delphi that should be considered by the corporate planner lie in yet unexplored areas. These are best framed in the form of questions. What is the psychological bias of the year 2000? (Are people afraid to name dates that cross the end of the second millenia? Does the year 2000 affect the perception of time, as dates are named that near it?) What individual characteristics, if common to a panel, influence its forecasts? Does the position of an individual in his career or life 'bias' results? What kind of perceived risk taking is involved in the minds of Delphi panelists? Is it possible that considering events in the future beyond a certain range is too 'inexact' a science and is not a legitimate part of planning? Do Delphi panels take full cognizance of the fact that their forecast may somehow influence the event? What are the confidence limits that can empirically be placed or justified for Delphi forecasts?

The answers to these and other questions raised by the preceding discussion are, certainly, not available readily. However, a consideration of their possible effects is paramount to the truly effective use of Delphi forecasting and they should, therefore, be resolved in some form by the corporate planner.

5. Recommendations for Corporate Utilization of Delphi

This article has presented material that contains various levels of implication for corporations desiring to apply the Delphi technique to forecasting. For purposes of clarity, the suggestions discussed below are limited to improving the basic structureⁱⁿ which Delphi is employed.

The planner or organization must first match the kind of output desired with what can be produced from Delphi. If this is not addressed and resolved at the outset, well before people are lulled into false security by sophisticated computer printout, the Delphi user may find himself (a) working with naive and intuitively obvious results, (b) promising more than can obviously be delivered, (c) making forecasts that are not used, or (d) helping to allocate strategy monies on an unsound basis. None of these

alternatives would appear desirable.

Next, the amount of tolerable uncertainty should be considered. It might be measured by assuming (not intuitively obvious) that it is analogous to the uncertainty of 'almanac' studies. An alternative and a problem by itself is the interquartile range, a measure of the dispersion of responses. If this range is five years when the median date is at 1975, seven years, when the median is 1980 and ten years at 1988, there may be a point in time where the range places limitations on the usefulness of the forecast. In other words, assessments of the utility of forecasts with various interquartile ranges (e. g. , a 1980 median forecast $\pm$ 5 years) should be made prior to establishing a Delphi based planning system.

Another area of interest is the question of utilizing Delphi alone or with other techniques. Since Delphi is a technique based on subjective modeling, it would be an advantage to integrate it with those that are more quantitative. This would provide links with historical data and help to yield a stronger forecast. In general, Delphi could be well integrated with trend analysis, mapping, and monitoring of the technological environment.

Two additional comments of use in the design of a Delphi forecast are directed at the consideration of the research planning or industry development cycle and the stratification of the forecast. The research cycle is important by supplying dates that will serve as targets upon which to focus a forecast. The industry development cycle can also provide helpful dates, but may be particularly important by suggesting aspects of technological developments that should be explored by the forecast, e. g. , capital funds constraints, sources and opportunities. Forecast stratification is the breakdown of all phases of the technology of interest into overlapping subgroups that may, in turn, be addressed by groups of self-rated experts. The overlapping will permit the development of internal consistency, while the use of subgroupings will yield greater accuracy.

Aside from technical matters, one of the advantages of Delphi is that it stimulates thinking and involves management in the forecasting process. This by itself could well be enough to justify its use.

In summary, then, a consideration of the behavior of Delphi combined with an awareness of its development should indicate the need for a certain amount of caution in applying the technique too vigorously.

Finally, it should also be apparent that Delphi is a very useful tool, which can be used to considerable advantage when applied with knowledge and forethought.

REFERENCES

- (1) Cetron, Marvin J. and Christine Ralph, Industrial Applications of Technological Forecasting, John Wiley Inc., New York, 1971.
- (2) Dalkey, **N.** and Olaf Helmer, "An Experimental Application of the Delphi Method to the Use of Experts", Management Science, TIMS, Baltimore, Md., 1963. v. 9, N. 3, April ↗
- (3) Kaplan, A., A. L. Skogstad, M. A. Girshick, "The Prediction of Social and Technological Events", Public Opinion Quarterly, Princeton, N. J., Winter, 1950.
- (4) Bernstein, G. B. and Marvin Cetron, "SEER: A Delphic Approach Applied to Information Processing", Technological Forecasting, American Elsevier Pub., V. 1, N. 1, June, 1969.
- (5) Gordon, T. J., "Forecasters Turn to Delphi", The Futurist, World Future Society, Bethesda, Md., V. 1, N. 1, February, 1967.
- (6) "Corporation Issues Major Forecast of Future Technology", The Futurist, World Future Society, Bethesda, Md., V. 1, N. 1, October, 1967.
- (7) Pyke, Donald L., "Mapping- A System Concept for Displaying Alternatives", Industrial Management Center, Austin Texas, 1970.
- (8) "Delphi Studies as an Aid to Corporate Planning", Industrial Management Center, Austin, Texas, 1970.
- (9) "Tire-Design Engineer Turns to Technological Forecasting", Product Engineering, December 21, 1970.
- (10) Dalkey, N. C., "Delphi", RAND Corp., Santa Monica, Calif., P-3704, October, 1967.

- (11) Gordon, T. J. and Robert H. Ament, "Forecasts of Some Technological and Scientific Developments and Their Societal Consequences", The Institute for the Future, Middletown, Conn., R-6, September, 1969.
- (12) Enzer, Selwyn, "A Case Study Using Forecasting as a Decisionmaking Aid", The Institute for the Future, Middletown, Conn., WP-2, December, 1969.
- (13) Gordon, T. J., "A Study of Potential Changes in Employee Benefits", The Institute for the Future, Middletown, Conn., V. 1, R-1, April, 1969.
- (14) " " V. 2, R-2 " .
- (15) " " V. 3, R-3 " .
- (16) Bregard, Raul de and Olaf Helmer, "Some Potential Societal Developments", The Institute for the Future, Middletown, Conn., R-7, April, 1970.
- (17) Gordon, T. J. and Olaf Helmer, "Report on a Long - range Forecasting Study" Appendix to ref. 26.
- (18) "McGraw Hill Survey of Technological Breakthroughs and Widespread Applications of Significant Technical Developments", Department of Economics, McGraw Hill Publications Company, 1968.
- (19) Parsons, P. and W. Williams, Forecast 1968-2000 of Computer Developments and Applications, Nyropsgade 47, 1602 Copenhagen V, Denmark, 1968.
- (20) Helmer, Olaf and Nicholas Rescher, "On the Epistemology of the Inexact Sciences", Management Sciences, TIMS, Baltimore, Md., Vol. 6, No.1, October 1959.
- (21) Lanford, HW., "A Synthesis of Technological Forecasting Methodologies", unpublished manuscript, Wright State University, Wright-Patterson AFB, Ohio, 1970.

- (22) Encyclopedia Brittanica, Vol. 7, "Delphi".
- (23) Brown, Bernice and Olaf Helmer, "Improving the Reliability of Estimates From a Consensus of Experts", RAND Corp., Santa Monica, Calif., P-2986, Sept. 1964.
- (24) Dalkey, N. "An Experimental Study of Group Opinion", Futures, Iliffe House Science and Technology Publications Ltd., Vol. 1, No. 5, Sept. 1969
- (25) Dalkey, N., "Analyses From a Group Opinion Study", Futures, Iliffe House Science and Technology Publications Ltd., Vol. 1, No. 6, Dec. 1969
- (26) Helmer, Olaf, Social Technology, Basic Books Inc., New York, 1966.
- (27) Bright, James R., ed., Technological Forecasting for Industry and Government, particularly, "The Delphi Method - An Illustration (Bright), "Technology, the Chicken- Corporate Goals, The Egg (Harper Q. North).
- (28) Gordon, T. J., "Cross-Impact Matrices", Future, Iliffe House Science and Technology Publications Ltd., Vol. 1, No. 6., Dec. 1969.
- (29) Gordon, T. J. and H. Hayward, "Initial Experiments with the Cross Impact Matrix Method of Forecasting", Future, Iliffe House Science and Technology Publications Ltd., Vol. 1, No. 2, Dec. 1968.
- (30) Kovac, Frederick J., "Technology Forecasting-Tires", Chemical Technology, Jan. 1971.
- (31) Ament, Robert H., "Comparison of Delphi Forecasting Studies in 1964 and 1969", Future, Iliffe House Science and Technology Publications Ltd., Vol. 2, No. 1, Mar. 1970.

- (32) Dalkey, N. B. Brown, S. Cochran, "Use of Self-Ratings to Improve Group Estimates", Technological Forecasting, American Elsevier Pub. Co., New York, Vol. 1, No. 3, Mar. 1970
- (33) Martino, Joseph P., "The Precision of Delphi Estimates", Technological Forecasting, American Elsevier Pub. Co., New York, Vol. 1, No. 3, March 1970.
- (34) Martino, Joseph P., "The Lognormality of Delphi Estimates", Technological Forecasting, American Elsevier Pub. Co., New York, Vol. 1, No. 4, Spring 1970.
- (35) Blackman, A. Wade, "The Use of Bayesian Techniques in Delphi Forecasts", Technological Forecasting, Vol. 2, No. 3/4, 1971.

Additional Sources;

Kovac, Frederick J. "Technology Forecasting- tires", Chemical Technology, Jan. 1971, P.18-23.

Landergren, Ulf; Stig Fredriksson, Ann Littke-Porsson, "An Application of the Delphi Technique to a Forecast of the Field of Ceramics- a Methodological Experiment," United Nations Economic Commission for Europe, Seminar on Technological Forecasting, Warsaw (Poland), 7-12 Dec. 1970.

"Institute for the Future Will Begin Operations in Connecticut this Fall" , The Futurist, World Future Society, Bethesda, Md., Vol. 2, No. 4, August 1968.

1000 1000

1000 1000

BACCHINI
Date Due

DEC 05 '75

APR 18 '76

May 15, '76

~~June 14, '76~~

FEB 3 '84

Lib-26-67

MIT LIBRARIES



3 9080 003 906 242

511-71

MIT LIBRARIES



3 9080 003 902 530

512-71

MIT LIBRARIES



3 9080 003 902 548

513-71

MIT LIBRARIES



3 9080 003 902 589

HD28
M17
514-71
No

"J. Line

MIT LIBRARIES



3 9080 003 875 272

515-71

MIT LIBRARIES



3 9080 003 869 341

516-71

MIT LIBRARIES



3 9080 003 875 249

517-71

MIT LIBRARIES



3 9080 003 869 325

518-71

MIT LIBRARIES



3 9080 003 875 223

519-71

MIT LIBRARIES



3 9080 003 875 231

520-71

MIT LIBRARIES



3 9080 003 906 135

521-71

MIT LIBRARIES



3 9080 003 875 181

522-71

